



Comparison of whole genome prediction accuracy across generations using parametric and semi parametric methods

Abbas Atefi, Abdol Ahad Shadparvar* and Navid Ghavi Hossein-Zadeh

Department of Animal Science, University of Guilan, Iran. *Author for correspondence: E-mail: shadparvar@yahoo.com

ABSTRACT. Accuracy of genomic prediction was compared using three parametric and semi parametric methods, including BayesA, Bayesian LASSO and Reproducing kernel Hilbert spaces regression under various levels of heritability (0.15, 0.3 and 0.45), different number of markers (500, 750 and 1000) and generation intervals of validating set. A historical population of 1000 individuals with equal sex ratio was simulated for 100 generations at constant size. It followed by 100 extra generations of gradually reducing size down to 500 individuals in generation 200. Individuals of generation 200 were mated randomly for 10 more generations applying litter size of 5 to expand the historical generation. Finally, 50 males and 500 females chosen from generation 210 were randomly mated to generate 10 more generations of recent population. Individuals born in generation 211 considered as the training set while the validation set was composed of individuals either from generations 213, 215 or 217. The genome comprised one chromosome of 100 cM length carrying 50 QTLs. There was no significant difference between accuracy of investigated methods ($p > 0.05$) but among three methods, the highest mean accuracy (0.659) was observed for BayesA. By increasing the heritability, the average genomic accuracy increased from 0.53 to 0.75 ($p < 0.05$). The number of SNPs affected the accuracy and accuracies increased as number of SNPs increased; therefore, the highest accuracy was for the case number of SNPs=1000. With getting away from validating set, the accuracies decreased and the most severe decay observed in the case of low heritability. Decreasing the accuracy across generations affected by marker density but was independent from investigated methods.

Keywords: accuracy, genomic, semi parametric methods, genetic architecture.

Comparação da acurácia de predições genômicas através de gerações usando métodos paramétricos e semi paramétricos

RESUMO. A acurácia da predição genômica foi comparada através de três métodos paramétricos e semi-paramétricos, que incluíram BayesA, LASSO Bayesiano e regressão RKHS (Reproducing Kernel Hilbert Spaces) sob vários níveis de hereditariedade (0,15; 0,3 e 0,45), números diferentes de marcadores (500, 750 e 1000) e intervalos de geração de conjuntos de validação. Uma população histórica de 1000 indivíduos com igual proporção sexual foi simulada por 1000 gerações, em tamanho constante. Ela foi seguida por 100 gerações adicionais com redução gradual de tamanho para 500 indivíduos na 200ª geração. Indivíduos da geração 200 foram cruzados aleatoriamente por mais 10 gerações aplicando-se tamanho de ninhada de 5, para expansão da geração histórica. Por fim, uma seleção de 50 machos e 500 fêmeas da geração 210 foram cruzados aleatoriamente, resultando em mais 10 gerações de população recente. Indivíduos nascidos na geração 211 foram considerados como o conjunto de treinamento, enquanto o conjunto de validação foi composto por indivíduos da geração 213, 215 ou 217. O genoma foi composto por um cromossomo de 100cM de comprimento portando 50 QTLs. Não houve diferença significativa entre a acurácia dos métodos investigados ($p > 0,05$); porém, dentre os três métodos, a maior média de acurácia (0,659) foi observada em BayesA. Com o aumento da hereditariedade, a média da acurácia de seleção genômica aumentou de 0,53 para 0,75 ($p < 0,05$). O número de SNPs afetou a acurácia, visto que seu valor aumentou com o aumento de SNPs; assim, a maior acurácia foi verificada para número de casos de SNPs=1000. Com o afastamento do conjunto de validação, a acurácia diminuiu e a redução mais pronunciada foi observada para o caso de baixa hereditariedade. A redução da acurácia através de gerações foi afetada pela densidade do marcador, mas foi independente dos métodos investigados.

Palavras-chave: acurácia, genômica, métodos semi paramétricos, arquitetura genética.

Introduction

Animal breeding aims to improve economic efficiency of livestock species through selection under a changing cost and income scenario. Most of the economic important traits in livestock have a

polygenic and quantitative expression, i.e., controlled by a large number of genes and affected by environmental factors. Statistical analysis of phenotypes and pedigree information allows prediction of the breeding values of the selection

candidates based on Fisher's infinitesimal model (Fisher, 1919). The best linear unbiased prediction (BLUP) (Henderson, 1974) has been extensively used for estimating the breeding values in animal and plant breeding programs and it leads to great progress in economic important traits.

Advances in the molecular genetics technologies have greatly increased the information available on individual genes for quantitative traits. Molecular techniques have enabled researchers to identify genetic markers that can be used to precise selection of candidates and thus increase efficiency of breeding programs. Genomic selection is a new tool for selection of animals based on marker information in addition to traditional information (Meuwissen, Hayes, & Goddard, 2001). The implementation of genomic selection takes place in two steps. Firstly, the effects of genetic markers, typically single nucleotide polymorphism (SNP), are estimated in a reference population (animals with both marker and phenotypic information). Secondly, the estimated effects of the genetic markers are used to predict genomic breeding values for individuals of validating population (animals with marker information and without phenotypic information). Meuwissen et al. (2001) suggested using genetic marker information in a statistical model of prediction of animal breeding values. They used three statistical models: a model assigning random effects to all available genetic markers ("genomic BLUP"), assuming that all markers effects are drawn from the same normal distribution, and two Bayesian models, where all ("BayesA") or a subset ("BayesB") of the random marker effects are drawn from distributions with different variances. Various types of these methods and additional methods have been subsequently suggested (Gianola, Campos, Hill, Manfredi, & Fernando, 2009). Gianola, Fernando, and Stella (2006) and Gianola and van Kaam (2008) have suggested a non-parametric treatment of genomic information by using Reproducing Kernel Hilbert Spaces (RKHS) regression, that subsequently have demonstrated with real data (González-Recio et al., 2008; González-Recio, Gianola, Rosa, Weigel, & Kranis, 2009).

Genomic selection provides a greater genetic progress in comparison with the traditional methods by increasing the accuracy of estimated breeding values and reducing generation intervals. The accuracy of genomic evaluation depends, among other factors, on the linkage disequilibrium (LD) between SNPs (Calus, De Roos, & Veerkamp, 2008).

LD is defined as the non-random association between the alleles at two different loci. LD can be caused by some factors including migration, mutation, selection or genetic drift in small populations, or any other event that may affect the genetic structure of the population. Population structure affects significantly the accuracy of genomic predictions when the selection candidates are closely related to the reference population (Habier, Fernando, & Dekkers, 2007; Habier, Tetens, Seefried, Lichtner, & Thaller, 2010) so that closer relatedness between subgroups (i.e., more recent divergence) increases LD between subgroups that leads to accurate genomic predictions (Daetwyler, Kemper, Van der Werf, & Hayes, 2012).

However, LD is decreasing due to recombination events in the meiosis before the development of the gametes of each new generation (Habier, Fernando, & Dekkers, 2009). To have sufficient reliability of genomic predictions, new genotyped and phenotyped individuals should be contributed in the reference population. Because of recombination events and decreasing the relatedness between subgroups, estimation of the marker effects should be re-evaluated at least every three generations (Hayes, 2007).

In this study, we investigated effect of marker density, trait heritability and decreasing the relatedness between training and validating population on prediction accuracy.

Material and methods

Simulation

The populations were simulated using the QMSim software (Sargolzaei & Schenkel, 2009) based on forward-in-time process. To achieve a mutation-drift equilibrium, a historical population consisted of 1000 unrelated animals (500 males and 500 females) were simulated 100 generation and then in order to create the linkage disequilibrium (LD), its size was gradually decreased from 1000 to 500 individuals in generation 200.

In the next step, individual of last historical generation (considered as founders) were mated randomly for 10 more generations, assuming five offsprings per dam and an exponential growth of the number of dams.

Finally, 50 males and 500 females from the last generation of the expanded population were randomly mated to generate another 10 generations. Individuals of generation 211 were considered as the training set and the individuals born at either generations of 213, 215 or 217 were regarded as validation sets.

The simulated genome consisted of one chromosome of 100 cM length. Aiming to investigate the effect of SNP density on prediction accuracy, 500, 750 or 1000 bi-allelic markers equally spaced across the genome were considered. Additive genetic effects were determined by 50 quantitative trait loci (QTL) randomly distributed through the genome. QTL effects were generated based on a normal distribution. The mutation rate of the markers and QTLs was assumed 2.5×10^{-5} per locus per generation (Solberg, Sonesson, & Woolliams, 2008).

A trait with heritability of 0.15, 0.30 and 0.45 was simulated. The true breeding value (TBV) of each individual was the sum of the QTL allele substitution effects and polygenic effects, assuming only additive effects. Phenotypes were generated by adding residuals, randomly drawn from a normal distribution with mean equal to zero, to the TBVs. The correlation between the true BV and the genomic predicted BV ($r_{TBV, GEBV}$) was used as a measure of the accuracy of GEBV prediction.

Linkage disequilibrium calculation

The extent of LD in the training populations was measured by r^2 :

$$r^2 = \frac{D^2}{freq(A1) * freq(A2) * freq(B1) * freq(B2)}$$

where $freq(A1)$ is the frequency of the A1 allele in the population, and likewise for the other alleles in the population and D is another statistic of linkage disequilibrium that calculated as:

$$D = freq(A1_B1) * freq(A2_B2) - freq(A1_B2) * freq(A2_B1).$$

PLINK software (Purcell et al., 2007) was used to calculate the LD.

Statistical analysis

The general structure for the models in linear form is:

$$y = \mu + \sum_{j=1}^p X_j g_j + e$$

where y is the vector of phenotypic records, μ is the overall mean, g_j is the coefficient of marker j denoting the allele substitution effect, X_j is a design matrix of genotype codes for the respective marker, and e is a vector of residuals. The data were analyzed using three different approaches: two Bayesian methods (BayesA and Bayesian LASSO) and one kernel based semi parametric method (RKHS),

which was introduced by Gianola et al. (2006) for genomic evaluation. The BGLR (Bayesian Generalized linear regression) package of R software (Pérez & Campos, 2014) was used to genomic evaluations.

BayesA

Meuwissen et al. (2001) proposed two hierarchical Bayesian models for GS denoted by BayesA and BayesB. In both methods data and variances of the marker positions need to be modeled. Inferences about model parameters are based on the posterior distribution. The BayesA approach applies the same prior distribution for all variances of the marker positions.

Bayesian LASSO

Park and Casella (2008) introduced the Bayesian LASSO (BL) method for estimating the regression coefficients. The Bayesian LASSO is also used in GS (Campos, Gianola, Rosa, Weigel, & Crossa, 2010; Campos et al., 2009; Long, Gianola, Rosa, & Weigel, 2011) using the hierarchical model with a likelihood function.

Reproducing kernel Hilbert spaces regression

Gianola et al. (2006) proposed a semi parametric method in the genomic evaluations as an alternative to SNPs regressions. The hope was that these methods are capable for capturing complex interaction patterns that may be difficult to account for in a linear model framework.

In this study, each scenario repeated 10 times and estimation of marker effects was performed using the R package, BGLR (Pérez & Campos, 2014). In order to investigate effects of heritability, marker effect estimation method, number of markers and the interval between training and validating sets and their interactions on genomic prediction accuracy, the PROC GLM of SAS software was used (SAS, 2004).

Results and discussion

Linkage disequilibrium

The r^2 values for different marker densities were presented in Table 1. As expected the r^2 value increased with increment of SNP density and highest value (0.13) was for SNP=1000.

In genome-wide association studies (GWAS) and genomic prediction studies; a minimum value of LD is needed to obtain accurate results. For instance, an average r^2 of 0.2 between markers and QTL is essential to detect QTL of moderate effect. The extent of genome-wide LD is largely determined by

marker density and the past effective population size. The marker-QTL LD phase breaks over generations due to recombination in the meiosis process in each generation. Therefore, the markers effect must reestimate at least every three generations (Hayes, 2007).

Table 1. The r^2 values for different marker numbers.

SNP number	500	750	1000
r^2 value	0.08	0.11	0.13

Prediction accuracy

The results of analysis of variance of accuracy were presented in Table 2. Main factors including heritability, number of markers, generation of validating set, and also interaction of $h^2 \times$ Number of markers, $h^2 \times$ generation of validating set and Number of markers \times generation of validating set, affected the accuracy significantly ($p < 0.05$).

Table 2. Output of Analysis of variance for accuracy.

Source of variation*	DF	SS	MS	F value	PR > F
Method	2	0.00044816	0.00022408	9.17	0.0022
H2	2	0.64965327	0.32482664	13298.9	<.0001
N_marker	2	0.19771151	0.09885576	4047.3	<.0001
Interval	2	0.03076771	0.01538385	629.84	<.0001
Method*h2	4	0.00033287	0.00008322	3.41	0.0338
Method*n_marker	4	0.00023608	0.00005902	2.42	0.0915
Method*interval	4	0.0001345	0.00003362	1.38	0.2861
H2*n_marker	4	0.06258846	0.01564712	640.62	<.0001
H2*interval	4	0.00391493	0.00097873	40.07	<.0001
N_marker*interval	4	0.00212542	0.00053135	21.75	<.0001
Method*h2*n marker	8	0.00029984	0.00003748	1.53	0.2215
Method*h2*interval	8	0.00025359	0.0000317	1.3	0.3118
Method*n_marker*interval	8	0.00011898	0.00001487	0.61	0.7579
H2*n_marker*interval	8	0.00628114	0.00078514	32.14	<.0001

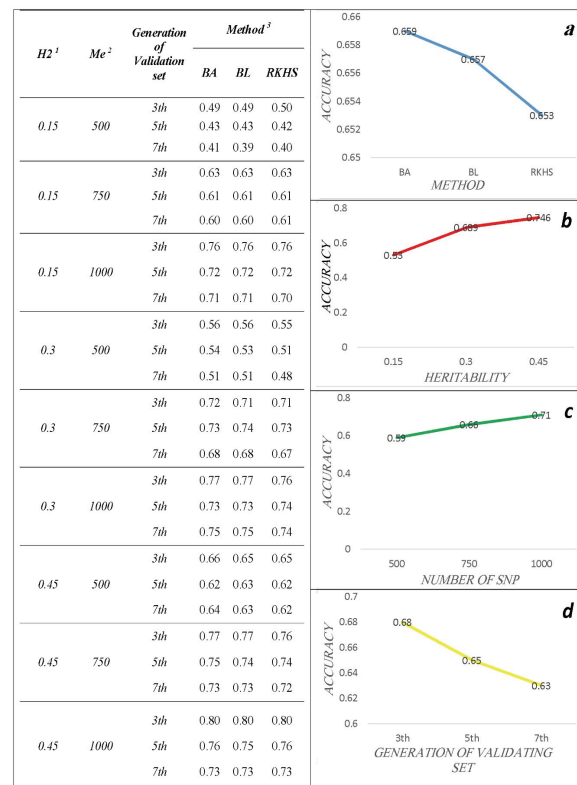
H2= heritability, N_marker= number of marker, Interval= the interval between training and validating sets.

Among investigated main factors, heritability and number of markers had the highest effect on accuracy. However, marker density is a factor easily controlled by researcher but heritability is a factor relating to trait structure and could not easily controlled by researcher.

Table 3 shows the accuracy of all investigated scenarios using three methods. Among three investigated methods, there were no differences in terms of accuracy but the highest mean accuracy $r_{TBV,GEBV} = 0.659$ was obtained in BayesA method (Figure 1a), while the computational demands of two other methods were higher. In RKHS method,

choosing the reproducing kernel and bandwidth value affected the results. In this study, the Gaussian reproducing kernel with the Euclidean distance between each pair of input vectors was chosen and the model using arbitrarily chosen bandwidth value was fitted. Gianola et al. (2006) reported that in comparison with RKHS and multiple linear regression (MLR) mixed model, when gene action was additive, these two methods had the same accuracy but when gene action was non additive (additive by additive interaction), the parametric MLR was clearly outperformed by RKHS. In a simulation study, the accuracies of BayesA and BL were the same and higher than RKHS (Howard, Carriquiry, & Beavis, 2014).

Table 3 and Figure 1 (a-d). Accuracy of different investigated scenarios (For all cases Standard Error < 0.04).



¹Heritability, ²Number of SNP, ³BayesA, Bayesian LASSO and Reproducing kernel Hilbert spaces regression.

With increasing the heritability, the correlation between TBV and GEBV was increased and the amount of excess was higher when departure was from 0.15 to 0.30 rather from 0.3 to .45 (Table 3 and Figure 1b). The high heritability equals with high contribution of gene effects in phenotypic variation and therefore the high accuracy. Many researchers have confirmed the high accuracy achieved for high heritabilities (Daetwyler, Villanueva, & Woolliams, 2008; Goddard, 2009; Howard et al., 2014). For high

heritability traits (such as carcass characteristics in meat animals), the gene effects have great contribution in phenotypic variation and therefore provide more accurate evaluations. Therefore, performing selection would easily lead to progress in this kind of traits.

The marker density is the most easily controllable factor by the researcher. The higher accuracy of genomic breeding values was associated to more dense markers (Figure 1c). Doubling the marker density, the accuracy increased from 0.59 to 0.71. The more marker density was led to more coverage of the genome and high LD which in turn resulted to better estimation of QTL effects. In a simulation study, doubling the number of SNP was led to increase the accuracy from 0.63 to 0.73 (Piyasatian & Dekkers, 2013) which is in agreement with our results. Increasing the accuracy of GEBV due to increment of marker study was reported in many studies (Combs & Bernardo, 2013; Piyasatian & Dekkers, 2013; Solberg et al., 2008)

Highest accuracy was observed when the interval between training and validating sets was as short as two generations (generation of validating set=3th). As time interval between training and validating sets prolonged, the accuracy of GEBV decreased. In a simulation study, accuracies were compared in the cases that validating set belonged to 1th and 7th generations and the accuracies were 0.73 and 0.60, respectively (Piyasatian & Dekkers, 2013). Accuracies decreased across generations because of the decay of genetic relationships that can captured by markers and changing of LD phase. In some researches, decreasing of genomic prediction accuracies over time due to breakdown of LD phase was reported (De Roos, Hayes, Spelman, & Goddard, 2008; Meuwissen et al., 2001). The recombination events break the LD phase during provision of gametes in each generation (Habier et al., 2009). In order to perform accurate genomic evaluations, marker effects should be re-evaluated at least every three generations (Hayes, 2007).

As mentioned above, the genomic accuracy increased along with increase in marker density, but the increment was different among three levels of heritability. At the low heritability (0.15), as the number of SNPs were doubled the accuracy increased 19 unites but at the high heritability (0.45), the excess only was 3 unites (Table 4). This suggested that dense SNP panels are more efficient in terms of accuracy, for traits with lower heritability.

The accuracy of genomic prediction decayed due to recombination effect across generations that change the LD phase between QTLs and markers.

The greater decline of accuracy was observed when the marker density was 500 and the decay of accuracy was decreased when the denser marker panels were used (Table 6). These results showed that using dense marker panels, the estimated marker effects of validating set could apply for longer time and it decreased the costs of re-estimating of marker effects. Habier et al. (2007) reported that accuracy of genomic prediction decreased across generations but opposite to our results the amount of decay depended on marker estimation method; therefore, the lowest and highest declines were for TP-BLUP and Bayes-B2, respectively.

Table 4. The estimated genomic accuracy for three marker density and three levels of heritability.

Heritability	Number of SNP		
	500	750	1000
0.15	0.44	0.53	0.63
0.30	0.61	0.71	0.74
0.45	0.73	0.75	0.76

As Table 5 shows, when heritability was 0.15, the accuracy of genomic evaluation for generation 7 of training set was 11 percent lower than the value for generation 3. The corresponding reductions for moderate and high heritabilities were 6 and 5 percent, respectively. As getting away from validating set, the accuracies decreased rapidly for low heritability case, suggesting that re-evaluating of marker effects is more imperative in the low heritability traits.

Table 5. The estimated genomic accuracy for three validating sets and three levels of heritability.

Heritability	Generation of training set		
	3 th	5 th	7 th
0.15	0.57	0.52	0.51
0.30	0.70	0.69	0.66
0.45	0.77	0.74	0.73

Table 6. The genomic accuracy for three validating sets and three levels of marker density.

Number of markers	Generation of training set		
	3 th	5 th	7 th
500	0.63	0.58	0.56
750	0.68	0.66	0.64
1000	0.74	0.71	0.70

Conclusion

The results of this study showed that for models with additive gene action RKHS method did not perform better than parametric methods such as BayesA and BL, although the RKHS is more complicated and time consuming. Comparison of

these methods for non-additive models should be performed under different simulation and real data. Marker density is one of the most important factor that affect the genomic prediction accuracy and fortunately by new progresses in genotyping technologies, the high density SNP panels with low cost is available and could employ easily for getting accurate genomic prediction. Preventing decay of accuracy due to recombination across time was one of most important benefits of dense marker panels so when the highest number of markers (1000) were used, the lowest accuracy decay was happened. In this study, the decreasing trend of accuracy across generations was not affected by marker effect estimation methods. In high heritability traits, increasing the number of markers had slight effect on accuracy but for low heritability trait, increasing number of markers increased accuracy severely; therefore, using the dense marker panels is imperative for low heritability traits. There was the same association between heritability and the interval between validating and testing sets, so that getting away from validating sets decreased the accuracy of high heritability trait slightly but the decline was severe for low heritability trait.

References

- Calus, M. P. L., De Roos, A. P. W., & Veerkamp, R. F. (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, *178*(1), 553-561.
- Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A., & Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research*, *92*(4), 295-308.
- Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... Cotes, J. M. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, *182*(1), 375-385.
- Combs, E., & Bernardo, R. (2013). Accuracy of genomewide selection for different traits with constant population size, heritability, and number of markers. *The Plant Genome*, *6*(1), 1-7.
- Daetwyler, H. D., Kemper, K. E., Van der Werf, J. H. J., & Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *Journal of Animal Science*, *90*(10), 3375-3384.
- Daetwyler, H. D., Villanueva, B., & Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, *3*(10), 3395.
- De Roos, A. P. W., Hayes, B. J., Spelman, R. J., & Goddard, M. E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, *179*(3), 1503-1512.
- Fisher, R. A. (1919). The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *52*(1), 347-363.
- Gianola, D., Campos, G., Hill, W. G., Manfredi, E., & Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics*, *183*(1), 347-363.
- Gianola, D., Fernando, R. L., & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, *173*(3), 1761-1776.
- Gianola, D., & van Kaam, J. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, *178*(4), 2289-2303.
- Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, *136*(2), 245-257.
- González-Recio, O., Gianola, D., Long, N., Weigel, K. A., Rosa, G. J., & Avendaño, S. (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics*, *178*(4), 2305-2313.
- González-Recio, O., Gianola, D., Rosa, G. J. M., Weigel, K. A., & Kranis, A. (2009). Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. *Genetics Selection Evolution*, *41*(1), 1-10.
- Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, *177*(4), 2389-2397.
- Habier, D., Fernando, R. L., & Dekkers, J. C. M. (2009). Genomic selection using low-density marker panels. *Genetics*, *182*(1), 343-353.
- Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P., & Thaller, G. (2010). The impact of genetic relationship information on genomic breeding values in German Holstein cattle. *Genetics Selection Evolution*, *42*(5), 2-12.
- Hayes, B. (2007). QTL mapping, MAS, and genomic selection. *A short-course. Animal Breeding & Genetics Department of Animal Science. Iowa State University*, *1*(1), 3-4.
- Henderson, C. R. (1974). General flexibility of linear model techniques for sire evaluation. *Journal of Dairy Science*, *57*(8), 963-972.
- Howard, R., Carriquiry, A. L., & Beavis, W. D. (2014). Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes | Genomes | Genetics*, *4*(6), 1027-1046.
- Long, N., Gianola, D., Rosa, G. J., & Weigel, K. A. (2011). Application of support vector regression to genome-assisted prediction of quantitative traits. *Theoretical and Applied Genetics*, *123*(7), 1065-1074.
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819-1829.

- Park, T., & Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482), 681-686.
- Pérez, P., & Campos, G. (2014). Genome-wide regression & prediction with the BGLR statistical package. *Genetics*, 1(1), 114.164442.
- Piyasatian, N., & Dekkers, J. (2013). Accuracy of genomic Prediction when accounting for population structure and polygenic effects. *Animal Industry Report*, 659(1), 68.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559-575.
- Sargolzaci, M., & Schenkel, F. S. (2009). QMSim: a large-scale genome simulator for livestock. *Bioinformatics*, 25(5), 680-681.
- Solberg, T. R., Sonesson, A. K., & Woolliams, J. A. (2008). Genomic selection using different marker types and densities. *Journal of Animal Science*, 86(10), 2447-2454.
- Statistical Analysis System [SAS]. (2004). *SAS/STAT User guide, Version 9.1.2*. Cary, NC: SAS Institute Inc.

Received on May, 22, 2016.

Accepted on July, 27, 2016.

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.